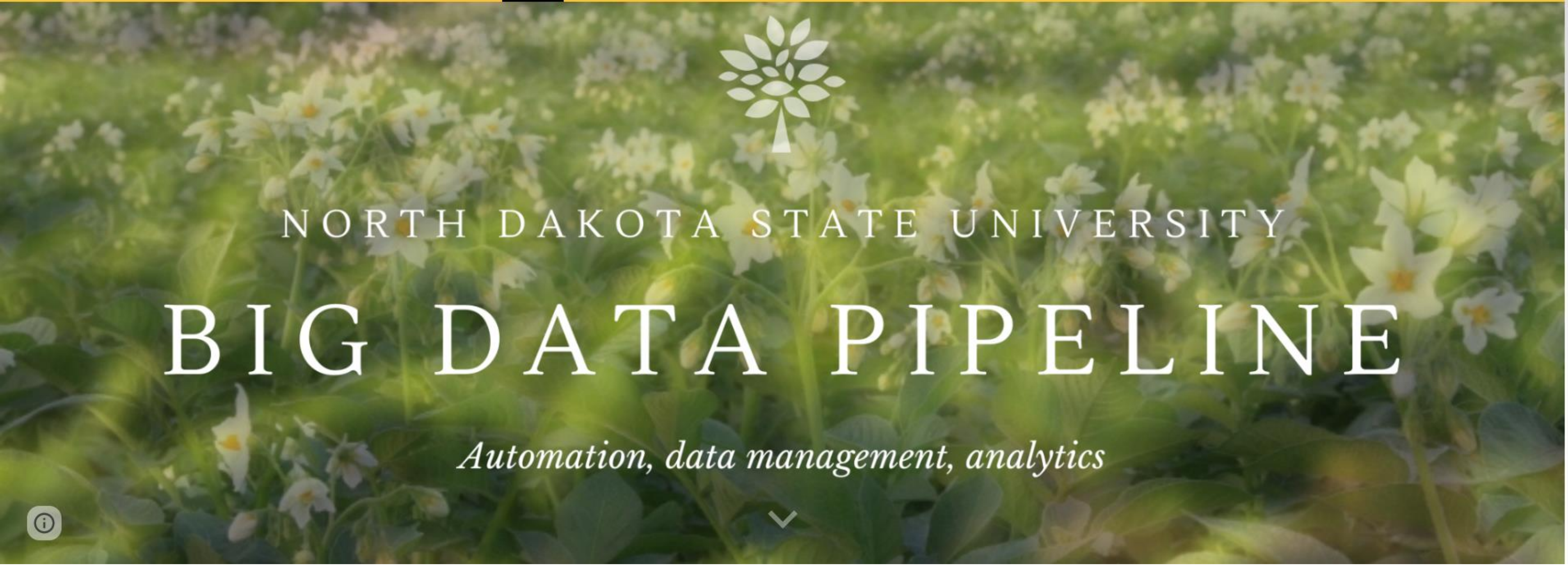
sites.google.com/ndsu.edu/plsc-bpdm/home

News | Find more details here

Big Data Pipeline Unit

Home | About us | Mission & Vision | Careers | Events & News | Projects | Collaborators | More



NDSU NORTH DAKOTA AGRICULTURAL EXPERIMENT STATION

40 Million acres

Avg farm size ~ 1,500 acres

Agriculture is
in our roots

42 Commodities

Generates $7.6 billion in cash receipts annual

Date:
Sept 29, 2022

NDSU Presenters

DIDIER MURILLO FLOREZ

Research Statistician Developer

Didier is a Mathematician and Statistician leading the development of statistical applications and algorithms that simplify end-user workflows at the ND AES Big Data Pipeline Unit. He is the main developer of FielDHub.

MATTHEW SEEFELDT

Software Development & Media Content Creator

Matthew is a Mathematician creating R-Shiny applications for our team. Additionally he develops multi-media projects for teaching tools for all our audiences. Main contribuitions made to FielDHub and Athena.

ANA MARIA HEILMAN

Senior Big Data Pipeline Manager

Ana is leading a growing team of data scientist, software developers, statisticians/mathematicians and bioinformaticist that support the ND Agricultural Experiment Station.

Main Contributors

RICHARD HORSLEY

Barley Breeder & Head
Department of Plant Sciences

Richard is a plant breeder with more than 40+ years of experience in plant breeding, experimental design, analysis of data, and academic administration. He is one of the minds behind the tools and technologies we develop in our team.

JOHAN APARICIO

Research Associate
Statistician, CIAT

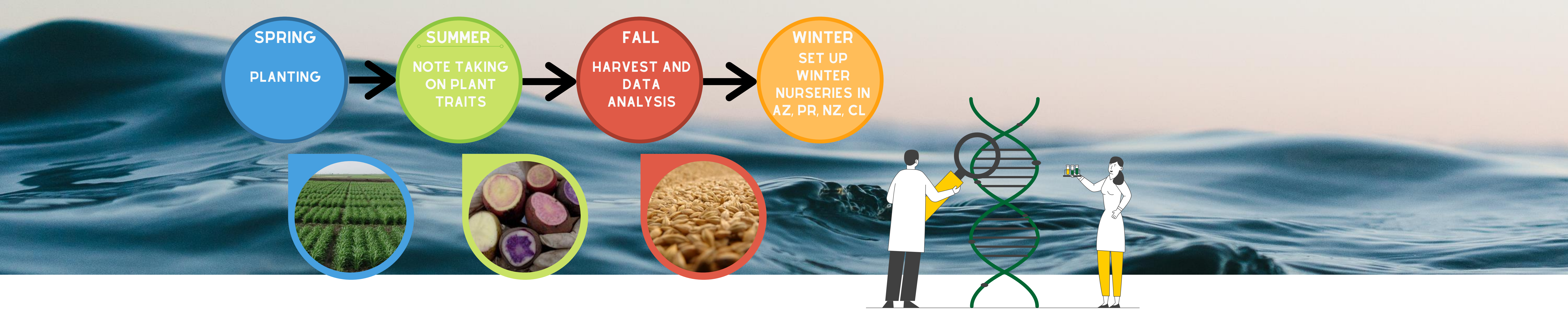
Johan is a Statistician leading the development of statistical applications and tools used by different research units at the International Center for Tropical Agriculture (CIAT). He is the main developer of Mr.Bean

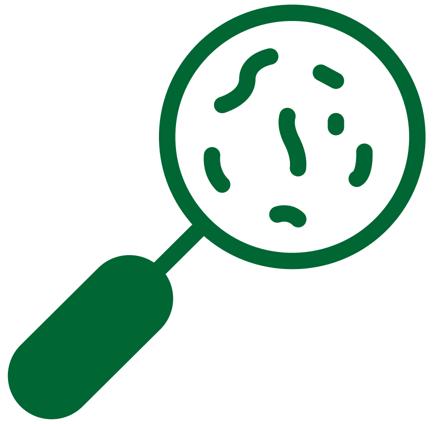
SALVADOR GEZAN

Statistical Genetics
Consultan, VSNi

Salvador is a Quantitative Geneticist and Plant Breeder, who is an Affiliate Faculty for the Department of Plant Sciences at NDSU. He is a contributor to both FielDHub and Mr.Bean applications, and also provides guidance to our team in analytic related projects.

Building Pipelines and Apps to Accelerate Discovery





Agriculture is
in our roots

Introduction to Design Of Experiments

Agriculture is
in our roots

What is Experimental Error?

5-10 word definition…

"Differences in experimental units treated alike"

An appropiate field experimental design should help you to:

1- Minimize Exp. Error

2- Measure the existing Exp. error

How to set up correctly your experiment?

1. Randomization
1. Replication
1. Blocking

Agriculture is
in our roots

Randomization is widely recognized as a basic principle of statistical experimentation

Randomization

- Reduces the bias
- Allows statistical tests
- Improves the accuracy in results

Agriculture is
in our roots

NDSU | NORTH DAKOTA AGRICULTURAL EXPERIMENT STATION

Replication

Increase the **precision**

Reduces **uncertainty**

Separates **background noise**

Agriculture is
in our roots

NDSU | NORTH DAKOTA AGRICULTURAL EXPERIMENT STATION

Blocking

Provides control of variability

Reduces error variances

BUT

Add only blocking factors that explain variation!

Agriculture is
in our roots

DOE in Agriculture and Plant Breeding

The origins of DOE comes from experiments in agriculture. There are a variety of applications in different areas such as soil science, weed science, plant breeding, and precision agriculture among many others.

➔Discard/select new genotypes in plant breeding programs
➔Compare fertilizer effects under different irrigations systems
➔Evaluate plant response under heat stress
➔Compare precision agriculture methods

Agriculture is
in our roots

Recap

01
Randomization

02
Replication

03
Blocking

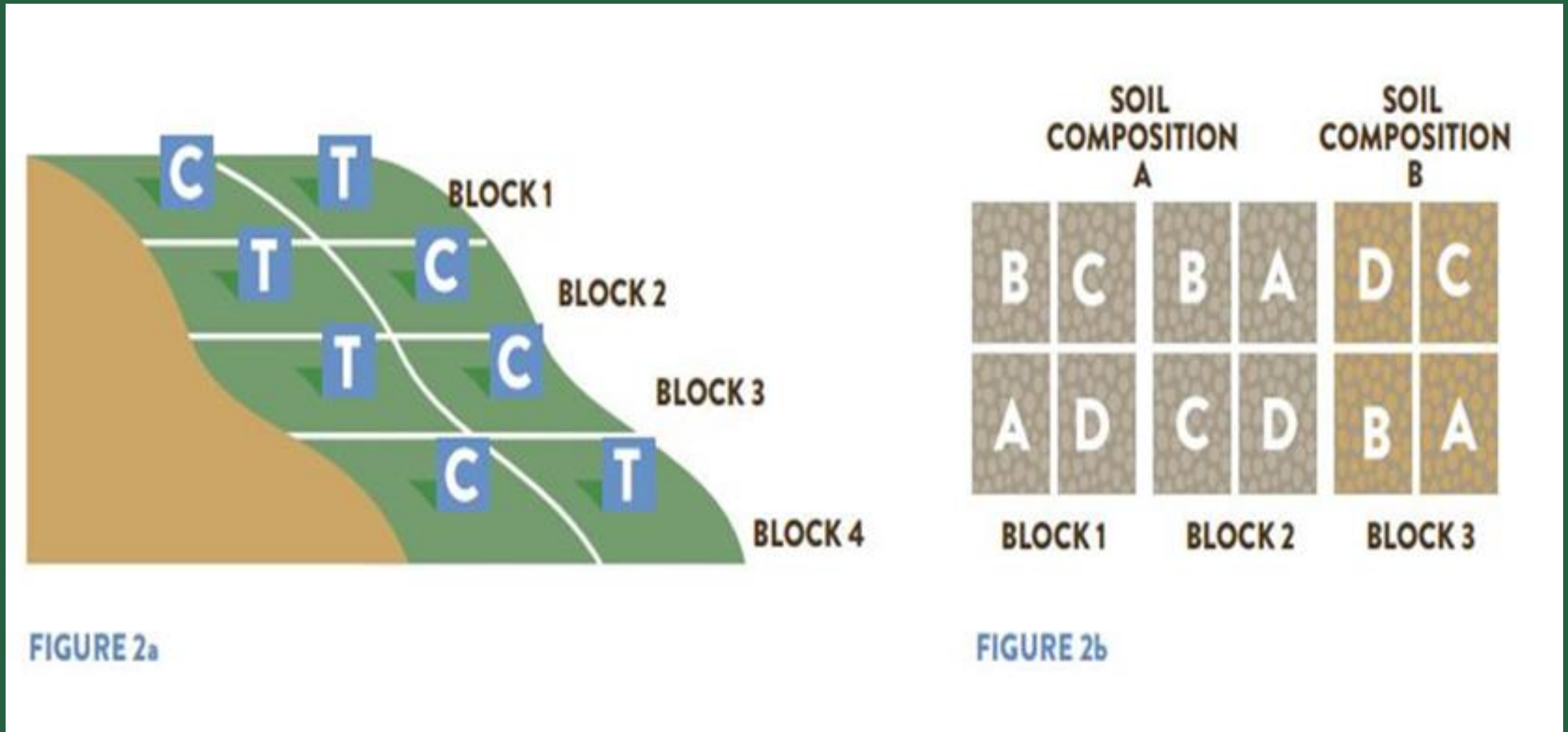


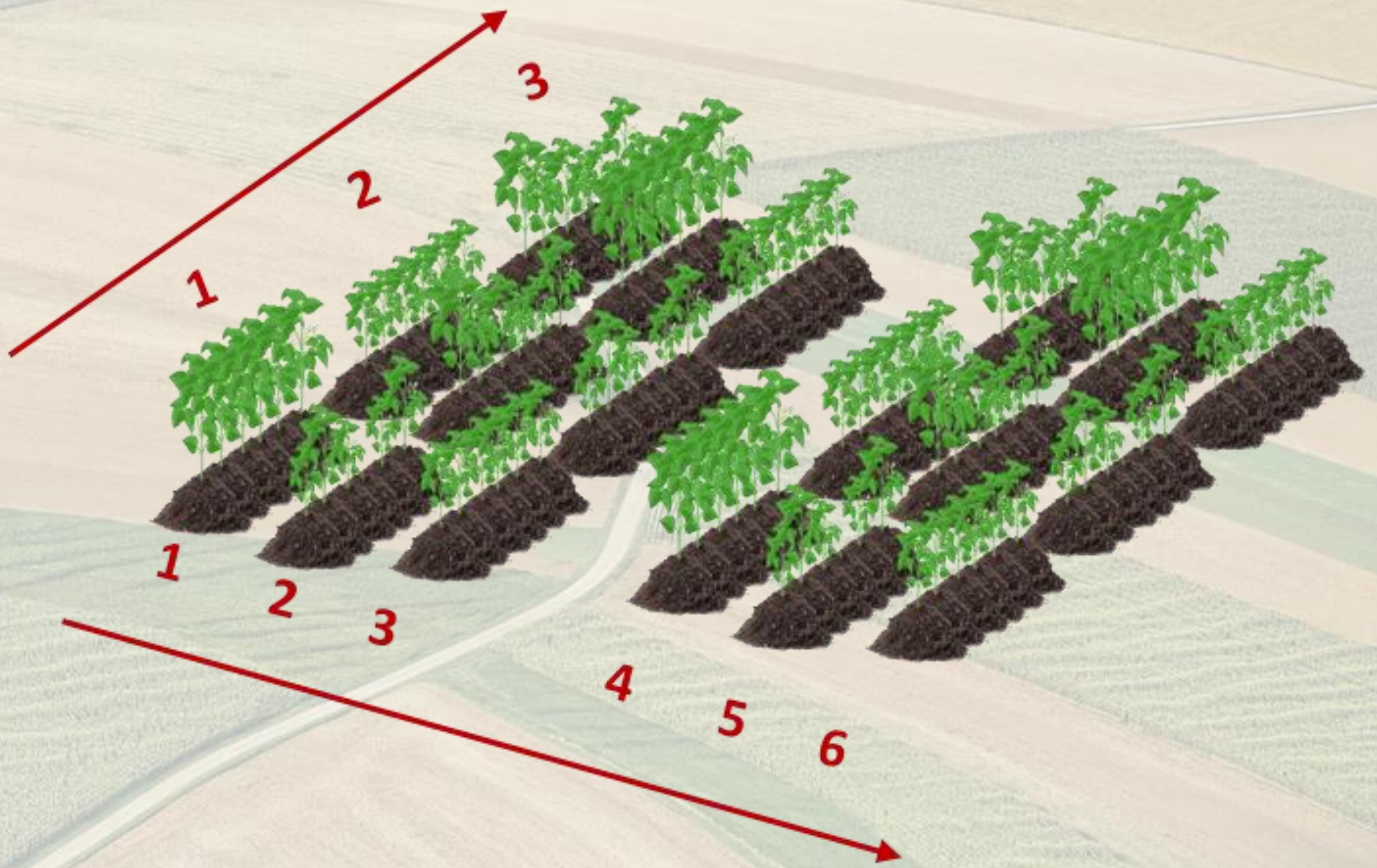
Figure from: https://www.sare.org/wp-content/uploads/how-to-conduct-research-on-your-farm-or-ranch.pdf

Agriculture is
in our roots

Field Heterogeneity

Spatial Correlation in field trials designs

- The experimental field is never completely homogeneous.
- Even before the plants are established in the field plots, there is some level of variability in soil properties.
- Experimental units that are closer together tend to be more similar than plots that are separated by greater distances.
- Row and column data are utilized to explain spatial variability in the field.



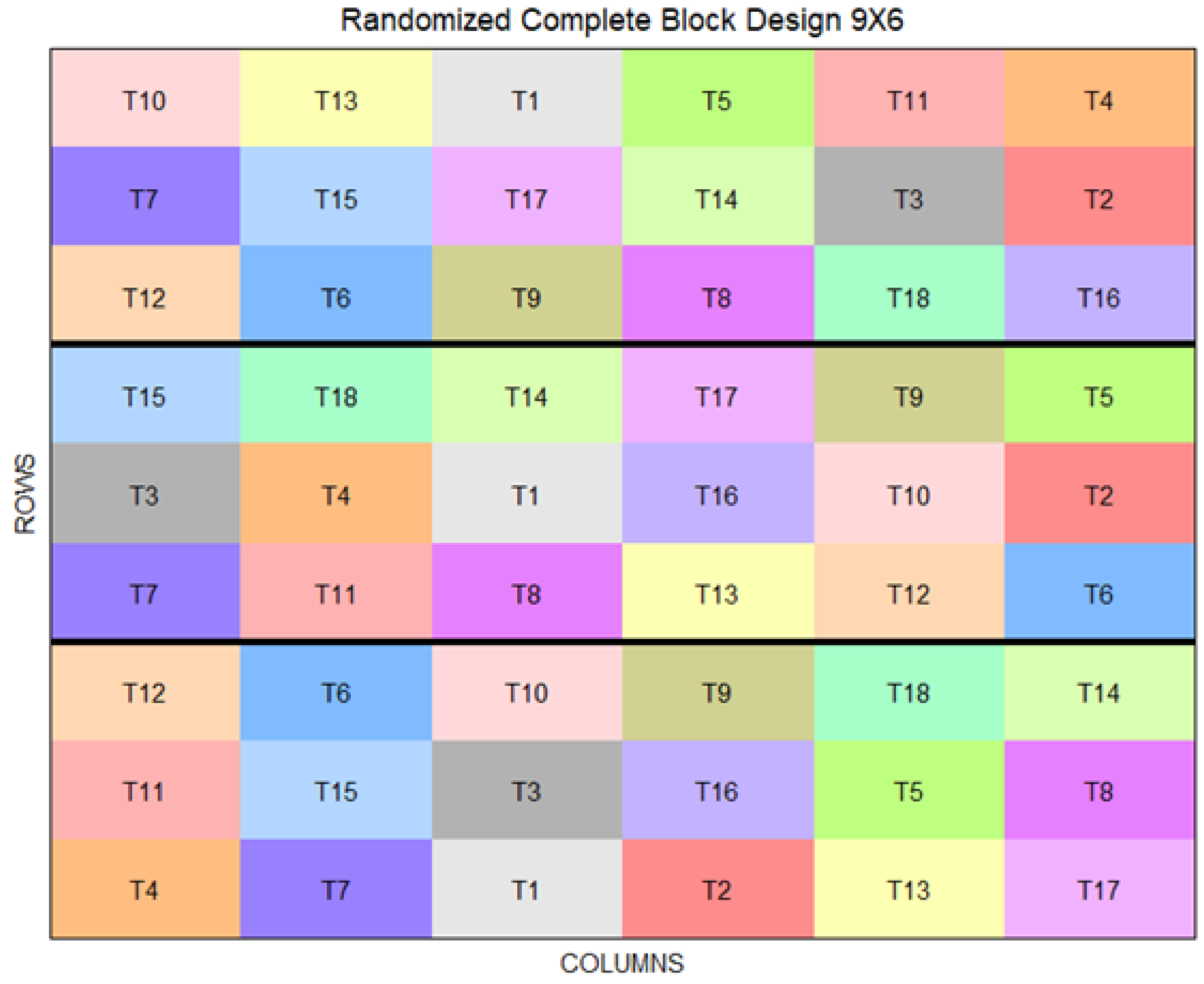
DOE

Experimental Designs in Agriculture (especially Plant Breeding)

- Randomized Complete Block Design (RCBD)
- Augmented RCBD
- Alpha-Lattice Design
- Partially Replicated Designs (prep)
- Unreplicated designs

Randomized Complete Block Design (RCBD)

- Each of the v treatments occurs once in every block (or replicate), and the number of units per block, k, is equal to the number of treatments (v=k).
- Advantage: balanced dataset, i.e. all treatment comparison has the same precision.
- Consider a study where we want to design an experiment with v = 18 treatments, where each will be replicated r = 3 times.



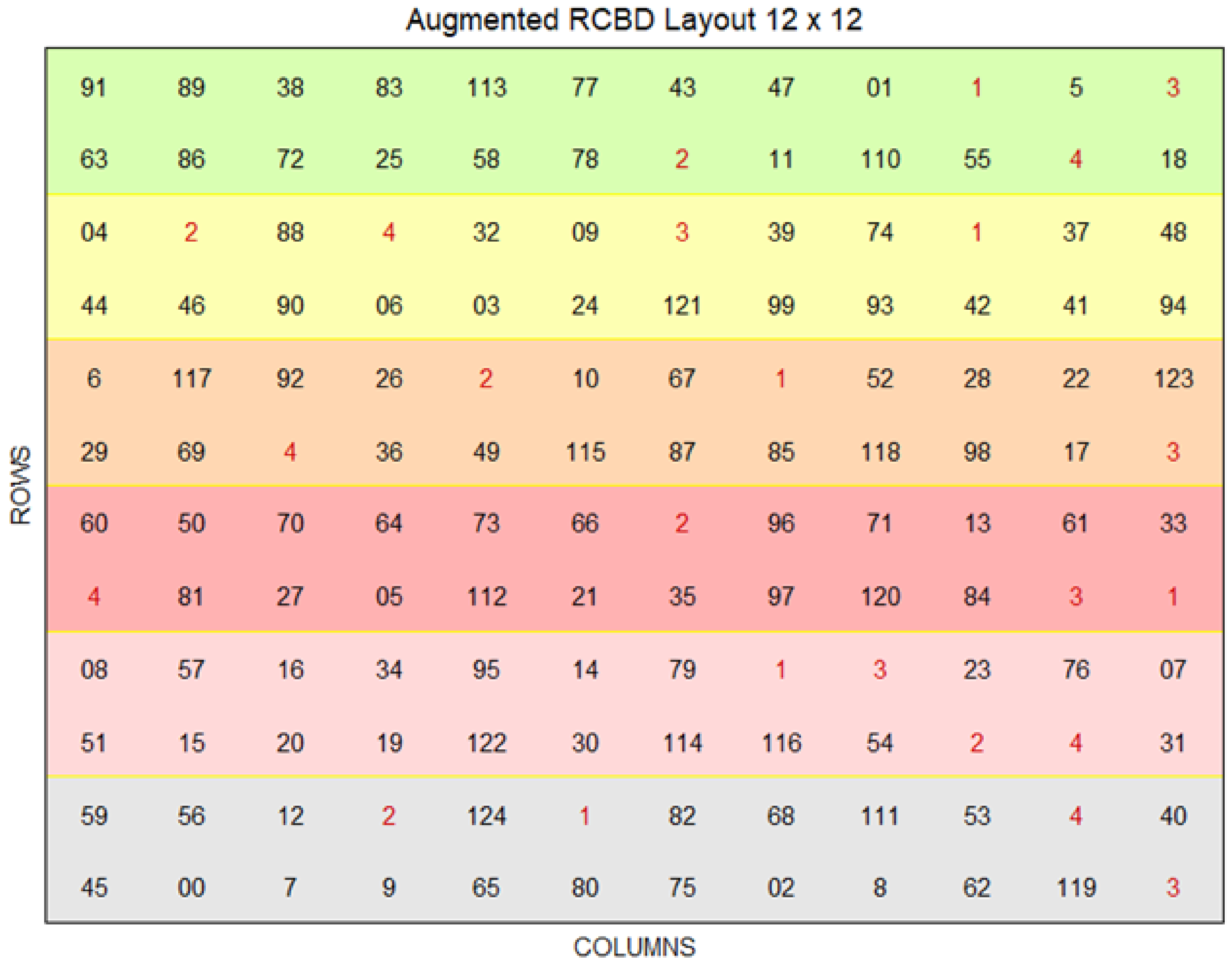
Agriculture is
in our roots

Augmented Design RCBD

- An augmented design is any standard design augmented with additional treatments in the complete block, incomplete block, row, column, etc. (Federer, 1961).

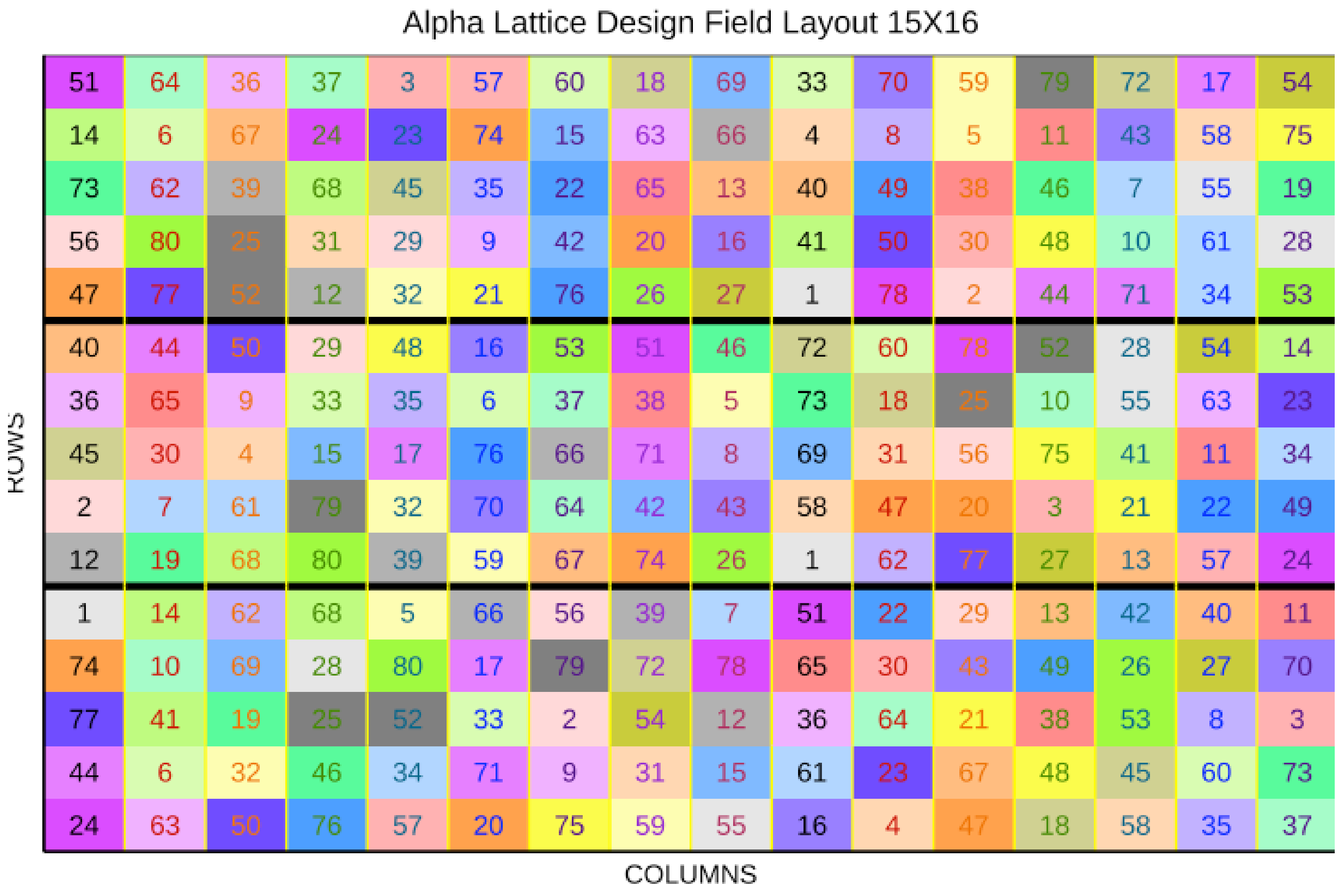
- Checks are analyses to estimate the error variance, which is later used for comparing the treatment levels.

- Consider an example with: 120 treatments, and 4 checks that are augmented to six blocks of size 2x12.



Alpha-Lattice Design

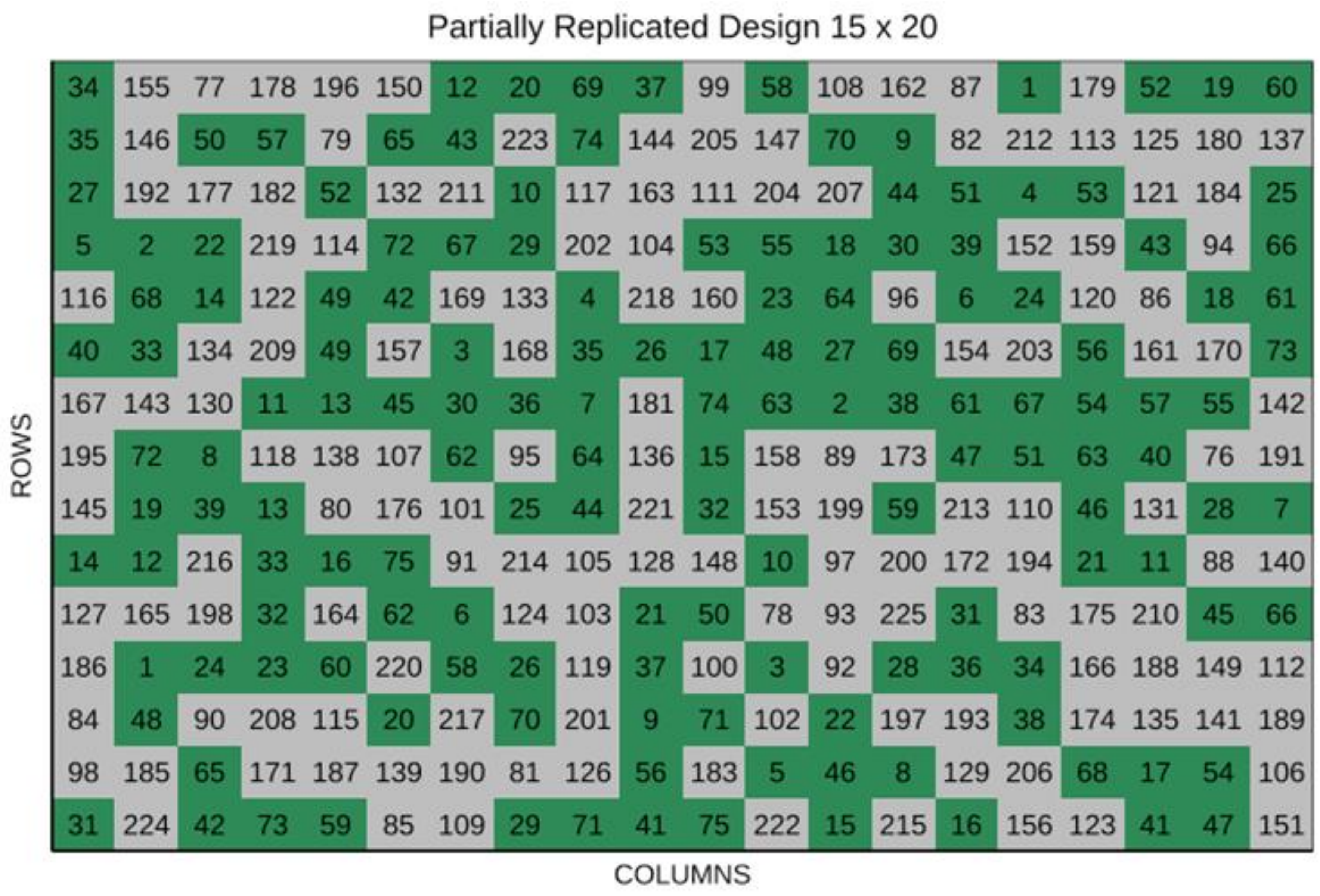
- When several treatments are evaluated, blocks can become quite large, and experimental units within the block are no longer homogeneous.
- Particularly recommended in cases where the number of treatments is large.
- Considering the example with 80 treatments, 3 reps and 5 units per incomplete block.



Partially Replicated Design

Designs where:

- A portion of treatments are replicated two or three times.
- The rest of the plots are considered for unreplicated treatments.
- Consider an example with: 225 treatments, 75 treatments appear twice and the rest 150 appear just one time.



Agriculture is
in our roots

Unreplicated Designs With Checks

Designs where:

- A large number of entries are unreplicated together with a group (usually 3 or 4) of check genotypes that are replicated in order to capture and model potential field effects.
- They enable the evaluation of a large number of genotypes (entries) in those cases with limited quantities of seed.
- Several arrangements exist: random, diagonal, optimized

Un-replicated Diagonal Arrangement 16 x 20

ROWS

247	2	167	229	76	79	22	46	205	2	206	31	230	16	176	62	126	3	278	152
149	270	253	181	2	130	7	114	244	173	17	81	2	137	106	49	277	84	127	273
191	128	15	261	164	156	103	1	5	141	190	63	284	240	269	1	174	25	160	222
162	215	1	28	235	27	40	9	165	255	4	138	45	148	35	276	219	111	3	119
64	256	30	281	26	3	267	24	10	258	187	82	60	4	68	109	32	90	239	50
4	204	87	271	144	225	196	241	1	36	245	142	192	33	131	43	2	20	200	89
42	100	237	1	212	280	139	199	231	182	73	4	184	168	56	178	155	117	124	2
195	77	234	66	198	57	3	107	242	260	248	185	217	145	3	18	147	12	140	61
179	4	169	58	54	83	194	65	14	1	37	272	47	150	268	251	122	1	158	283
104	80	101	249	4	102	41	59	44	186	38	39	3	105	125	226	75	197	69	108
218	211	29	74	282	34	171	1	202	115	151	8	193	243	210	2	159	264	180	236
129	153	3	110	95	203	170	161	67	85	3	11	92	177	189	172	232	88	1	98
209	123	86	224	52	4	201	275	99	157	94	55	254	2	207	19	53	6	257	143
4	48	112	96	259	279	113	21	4	252	223	71	274	146	175	116	2	216	51	78
233	166	188	2	134	227	135	263	133	120	136	3	228	250	154	132	70	221	118	3
266	93	163	121	91	213	4	72	13	214	265	208	23	220	1	246	262	238	97	183

COLUMNS

NDSU NORTH DAKOTA AGRICULTURAL EXPERIMENT STATION

Multiomic analyses in plant breeding

Designing & Analyzing Your Experiments Using FielDHub & Mr.Bean

Presented by Big Data Pipeline Team
Didier Murillo Florez
Matthew Seefeldt
Johan Aparicio
Ana Maria Heilman

Part 3:

FielDHub

Agriculture is
in our roots

Date:
Sept 29, 2022

What is FielDHub

- FielDHub is an R package/Shiny app that allows creating randomizations for a variety of experimental designs with applications in agriculture, plant breeding, forestry, animal, and biological sciences.

- The Big Data Pipeline Unit at NDSU is in charge of the development and maintenance of FielDHub. The app release was on May 2021 at CRAN. Recently, the team released a second version with new features.

FielDHub Sources

GitHub: https://github.com/DidierMurilloF/FielDHub

CRAN: https://cran.r-project.org/web/packages/FielDHub/index.html

Web Page: https://didiermurillof.github.io/FielDHub/

JOSS Paper: https://joss.theoj.org/papers/10.21105/joss.03122

YouTube: https://www.youtube.com/watch?v=Q9bftid0kPw&t=75s

Questions

Alliance

Bioversity International

CIAT
International Center for Tropical Agriculture
Since 1967 Science to cultivate change

BIG DATA
PIPELINE TEAM | NDSU

NDSU NORTH DAKOTA AGRICULTURAL EXPERIMENT STATION

Thank you

Agriculture is
in our roots